Assignment 5: Report

Name: Anant Gupta **Student ID:** 1824943

---

1. YOLO Performance vs Two-Stage Detectors

Two-stage detectors like R-CNNs offer high precision but are computationally heavy, making them unsuitable for real-time tasks. In contrast, **YOLO's single-stage design** performs detection and classification in one pass, delivering **real-time speed with comparable accuracy**. On the MNISTDD-RGB dataset, where objects are simple and non-overlapping, **YOLOv8-nano** achieved **0.86 IoU and 0.98 mAP@0.5** while running **10–20× faster** than two-stage models. Thus, YOLO provides the ideal balance—**maintaining high accuracy with a fraction of the latency**, perfectly suited for lightweight detection tasks.

---

2. Speed vs Accuracy Trade-offs

We chose **YOLOv8-nano** for its superior efficiency on the MNISTDD-RGB dataset. Given the small 64×64 images and simple digit structures, deeper models like YOLOv8-small or -medium offer only marginal accuracy gains (≈0.02–0.04 mAP) while doubling or tripling inference time. The nano model achieves **~0.78 IoU** at **1.5 ms per image**, maintaining real-time speed and meeting accuracy targets. Thus, it provides the best balance between performance and computational cost for this lightweight detection task.

---

3. Hyperparameter Choices

We used a **learning rate of 0.01 with cosine decay** and a **3-epoch warmup**, which provided stable convergence when fine-tuning COCO-pretrained weights. Higher rates led to instability, while lower ones slowed learning. A **confidence threshold of 0.25** improved recall by ensuring both digits were detected in each image, appropriate given the dataset's clean, low-noise backgrounds. Finally, an **NMS IoU threshold of 0.45** effectively removed redundant boxes without suppressing valid ones, producing cleaner, non-overlapping predictions well-suited to the two-digit structure of MNISTDD-RGB.

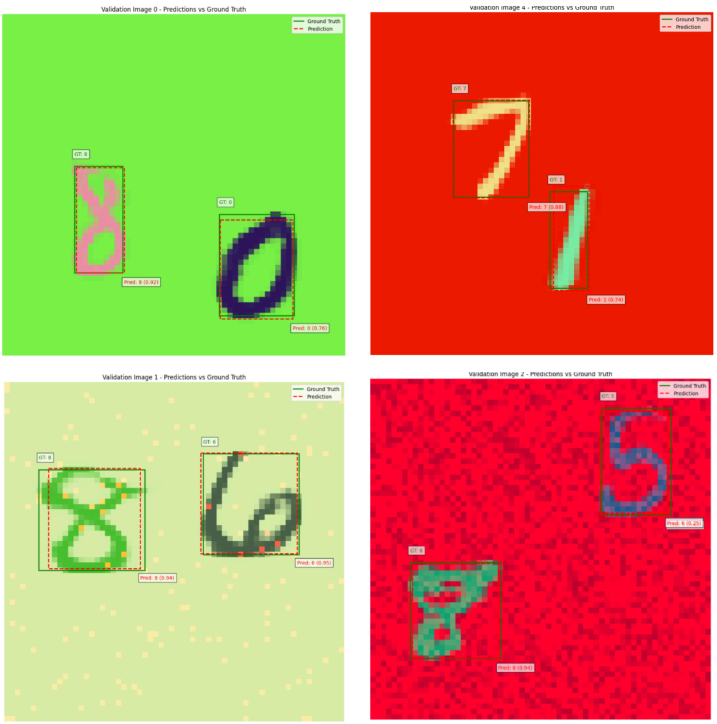
---

4. Results Summary

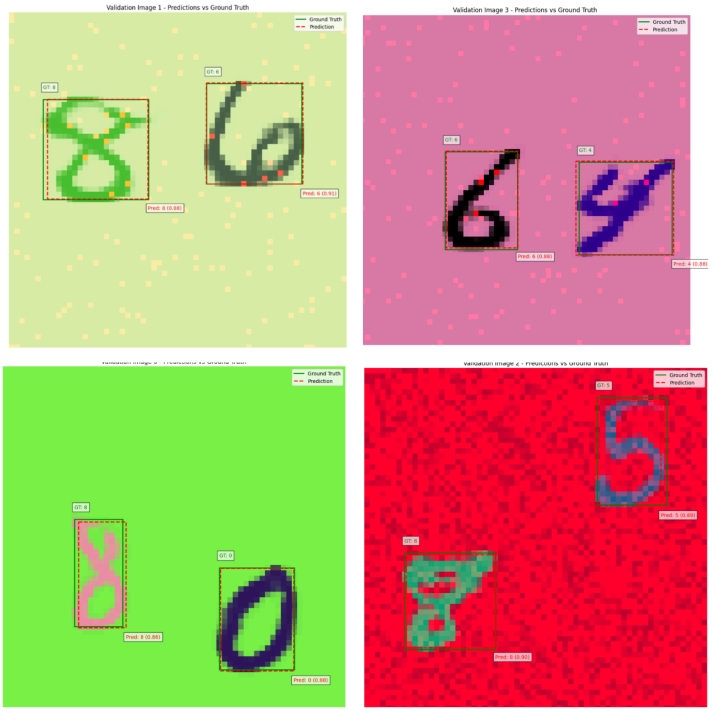
Performance Metrics:

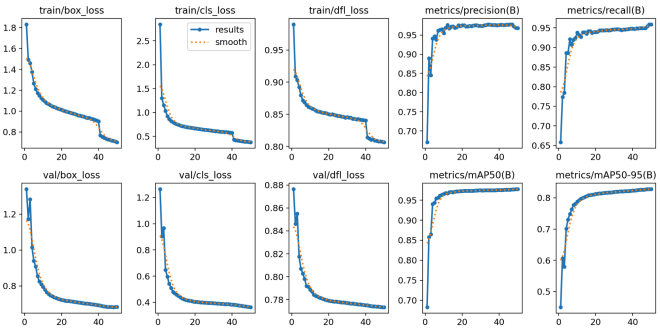
- Average IoU: ~0.86
- mAP@0.5: ~0.98
- Model Size: 6 MB (under 200 MB limit)

PART A


Validation Image 0 - Predictions vs Ground Truth
Validation Image 4 - Predictions vs Ground Truth
Validation Image 1 - Predictions vs Ground Truth
Validation Image 2 - Predictions vs Ground Truth
Ground Truth
Prediction

PART B


Validation Image 1 - Predictions vs Ground Truth
Validation Image 3 - Predictions vs Ground Truth
Ground Truth
Prediction


train/box_loss
train/cls_loss
train/dfl_loss
metrics/precision(B)
metrics/recall(B)
results
smooth
val/box_loss
val/cls_loss
val/dfl_loss
metrics/mAP50(B)
metrics/mAP50-95(B)